

Розробка моделі сегментації зображень з використанням згорткової нейронної мережі

Б. П. Книш, Я. А. Кулик

Розглянуто модель сегментації зображень за допомогою згорткових нейронних мереж та дослідження ефективності процесу на основі моделей із навчанням глибоких шарів згорткових нейромереж. Існують об'єктивні труднощі, пов'язані з визначенням оптимальних характеристик нейронних мереж, тому є проблема перенавчання нейромережі. Усунення перенавчання шляхом визначення лише оптимальної кількості епох недостатнє, оскільки не забезпечує високу точність.

Визначено вимоги до набору зображень для навчання та перевірки моделі. Даним вимогам найбільш відповідають набори зображень PASCAL VOC (Велика Британія) і NVIDIA-Aerial Drone (США).

Встановлено, що AlexNet (Канада) є навченою моделлю і може виконувати сегментацію зображень, проте надійність розпізнавання об'єктів недостатня. Тому виникає необхідність підвищення ефективності сегментації зображень. Доцільно використати архітектуру AlexNet для створення спеціалізованої моделі, яка за рахунок зміни параметрів та перенавчання деяких шарів дозволить краще проводити процес сегментації зображення.

Виконано навчання п'яти моделей з використанням таких параметрів: швидкість навчання, число епох, алгоритм оптимізації, вид зміни швидкості навчання, коефіцієнт гатта, попередньо навчена модель.

Розроблено згорткову нейронну мережу для підвищення точності та ефективності сегментації зображення. Визначено оптимальні параметри навчання нейромережі: швидкість навчання – 0,0001, число епох – 50, коефіцієнт гатта – 0,1 тощо. Отримано підвищення точності на 3 %, яке дозволяє стверджувати про правильність вибору архітектури розробленої мережі та підбору параметрів. Це дає можливість використовувати дану мережу для практичних задач сегментації зображень, зокрема для пристроїв із обмеженими обчислювальними ресурсами.

Ключові слова: обробка зображень, сегментація зображень, згорткові нейронні мережі, безпілотний літальний апарат.

1. Вступ

Обробка зображень має надзвичайно важливе значення в сучасній науці та практиці, тому постійно розвивається та вдосконалюється. Обробку зображень можна використовувати в багатьох галузях виробництва, а саме в точному землеробстві (агромоніторингу), системах безпеки, контролі якості тощо. У перелічених галузях задіяні системи технічного зору, робототехнічні комплекси, без-

пілотні літальні апарати (БПЛА), системи відеоспостереження, веб-сервіси та мобільні застосування для ідентифікації та пошуку.

Одним з видів обробки зображення є сегментація, яка широко застосовується в промисловості, мистецтві, медицині, космосі, при керуванні процесами, автоматизації та в багатьох інших галузях [1]. Сегментація зображень передбачає розбиття вхідного зображення на області, що не подібні по деякому критерію. Результатом цього є множина областей, які всі разом покривають все вхідне зображення.

Існує велика кількість методів сегментації зображень, серед яких найбільш поширеними є методи на основі аналізу графа зображення, кластеризації, контурні та порогові методи, а також нейромережеві методи. При цьому ці методи [2–4] працюють набагато швидше при невеликих розмірах зображень та кольоровій гаммі.

Оскільки сегментація передувє більш високому рівню обробки зображення, то до методів сегментації застосовуються певні вимоги. У загальному випадку ці вимоги можна сформулювати наступним чином:

- максимальна відповідність сегментованої області реальному об'єкту;
- висока швидкодія;
- стійкість до помилок;
- висока точність.

Тому виникає необхідність аналізу методів сегментації зображень та вибору оптимального згідно вищенаведених вимог, зокрема високої точності. Також варто врахувати параметри, які характеризують ці методи, зміна яких здійснює безпосередній вплив на точність, швидкодію та загальну ефективність процесу сегментації.

Сучасним актуальним напрямком виробництва є розвиток точного землеробства, яке ґрунтується на результатах агромоніторингу, а саме на знімках із відеокамер БПЛА для аналізу вегетативності, оцінки площ пошкоджень посівів, прогнозування врожайності тощо. Важливим критерієм при цьому виступає здатність БПЛА уникати зіткнень із близькими предметами, визначати положення в просторі, напрям та траєкторію польоту за рахунок надходження вхідних даних у вигляді сегментованих зображень.

Ефективність цих систем визначається точністю сегментації, для оцінки якої необхідні експериментальні дослідження.

2. Аналіз літературних даних та постановка проблеми

В роботі [5] наводяться результати сегментації зображень за допомогою кластеризації. Показано пошук кластерів на зображеннях на основі схожих характеристик. Даний процес характеризується високою швидкістю, точністю та стійкістю до помилок на початковому етапі. Але залишаються невирішені питання, пов'язані з суттєвим зменшенням ефективності методу при збільшенні розміру зображення. Причиною цього можуть бути об'єктивні труднощі, пов'язані з великою кількістю дрібних елементів на зображенні та поганою роботою методу на зашумленому зображенні, що робить відповідні дослідження недоцільними. В роботі [6] розглядаються порогові методи сегментації зобра-

жень, які характеризуються високою швидкістю та простотою в реалізації. Проте залишаються невирішені питання, пов'язані з доволі низькою точністю сегментації. Це викликає труднощі, пов'язані з необхідністю обмежувати палітру кольорів на границях елементів зображень, що обмежує використання методу для повноколірних зображень. В роботі [7] наводяться результати сегментації зображень за допомогою контурних методів. Показано стійкість до зміни параметрів вхідних зображень. Але залишається невирішеною проблема розриву областей зображення. Також методи характеризуються низькою швидкістю. Причиною цього можуть бути об'єктивні труднощі, пов'язані з використанням широкої палітри кольорів у реальних зображеннях. Це обмежує використання даного методу для фотореалістичних зображень. В роботі [8] наводяться результати сегментації зображень за допомогою методів на основі аналізу графа зображення. Показано, що отримані дані не залежать від зміни параметрів однорідності кольорів і розмірів вхідних зображень. Проте дані методи мають низьку швидкість та потребують багато пам'яті. Причиною цього можуть бути складнощі вибору вибирати метрику окремо для кожного типу зображень та використанням великої кількості елементів графу для сегментації зображення. Це робить дані дослідження затратними з точки зору використання обчислювальних ресурсів.

Варіантом подолання вищенаведених труднощів, пов'язаних з недостатньою точністю, ефективністю та швидкістю, може бути використання методів сегментації зображень на основі нейронних мереж [9], зокрема згорткових. Саме такий підхід використаний у роботі [10], де використовується нейронні мережі для кольорових зображень, але не аналізуються параметри навчання. Також схожий принцип реалізований в роботі [11], де параметри навчання наводяться без пояснення можливостей використання, зокрема для задач сегментації.

Все це дозволяє стверджувати, що доцільним є проведення дослідження по підвищенню ефективності навчання нейронної мережі, що дозволить значно покращити точність сегментації зображень.

3. Мета та задачі дослідження

Метою дослідження є вдосконалення архітектури згорткової нейронної мережі для сегментації зображень та вибору параметрів навчання цієї мережі. Це дасть можливість отримати нову нейромережу із підвищеною точністю для сегментації зображень, яка може бути використана як попередньо навчена нейромережа для інших задач.

Для досягнення мети були поставлені наступні завдання:

- дослідити моделі нейромереж на основі набору PASCAL VOC;
- оцінити модель Voc-3 на основі набору NVIDIA-Aerial Drone.

4. Матеріали та методи дослідження сегментації зображень за допомогою згорткових нейронних мереж

Дослідження сегментації зображень проводилось за допомогою відповідних методів на основі згорткових нейронних мереж з врахуванням параметрів навчання нейромережі. Для перевірки ефективності цих методів використову-

вались набори PASCAL VOC та NVIDIA-Aerial Drone, які містять велику кількість зображень із розміченими групами пікселів та визначеними класами об'єктів. PASCAL VOC містить зображення, де потрібний клас чітко виділяється з поміж інших пікселів по колірному тону [12]. NVIDIA-Aerial Drone містить зображення, які отримані з відеокамер, прикріплених на БПЛА при зйомці з висоти декількох сотень метрів [13]. Дослідження проводилось за допомогою програмного середовища DIGITS з використанням середовища Caffe, яке призначено для глибокого навчання нейромережі з врахуванням швидкодії та модульності при розробці моделі. Поєднання цих середовищ дозволяє швидко навчати нейронні мережі з глибокими шарами та використовується для задач класифікації, сегментації зображень та виявлення об'єктів на них. DIGITS містить попередньо навчену модель AlexNet, яка характеризується адаптованими для сегментації параметрами (табл. 1) та має гнучку архітектуру (рис. 1).

Таблиця 1
Параметри моделі AlexNet

Рівень	Розмір ядра	Розмір фільтра	Крок	Заповнення	Розмір даних	Функція активації
Вхід Input	—	—	—	—	227×227×3	—
Шар Conv1	96	11×11	4	—	55×55×96	ReLU
Шар Max Pool1	—	3×3	2	—	27×27×96	—
Шар Conv2	256	5×5	1	2	27×27×256	ReLU
Шар Max Pool2	—	3×3	2	—	13×13×256	—
Шар Conv3	384	3×3	1	1	227×13×384	ReLU
Шар Conv4	384	3×3	1	1	13×13×384	ReLU
Шар Conv5	256	3×3	1	1	13×13×256	ReLU
Шар Max Pool3	—	3×3	2	—	6×6×256	—
Вихід Dropout1	—	—	—	—	6×6×256	Softmax

Архітектура AlexNet складається з п'яти згорткових шарів, між якими розташовані шари Pooling і нормалізації, та трьох повнозв'язних шарів. Причому їх параметри можуть змінюватись в процесі навчання. На вхід подається зображення з палітрою RGB розміром 227×227×3. Розмір фільтра першого шару становить 11×11. Використовується ядро розміром 96 з кроком 4. Функція активації на виході – Softmax, а в шарах – ReLU, яка дозволяє в 6 разів збільшити швидкодію. Загалом кількість параметрів AlexNet становить 62,3 мільйона [14]. В порівнянні з аналогічними моделями [10], до AlexNet висуваються значно менші вимоги до величини пам'яті (в 10 раз) при підвищенні точності більш, ніж на 70 %.

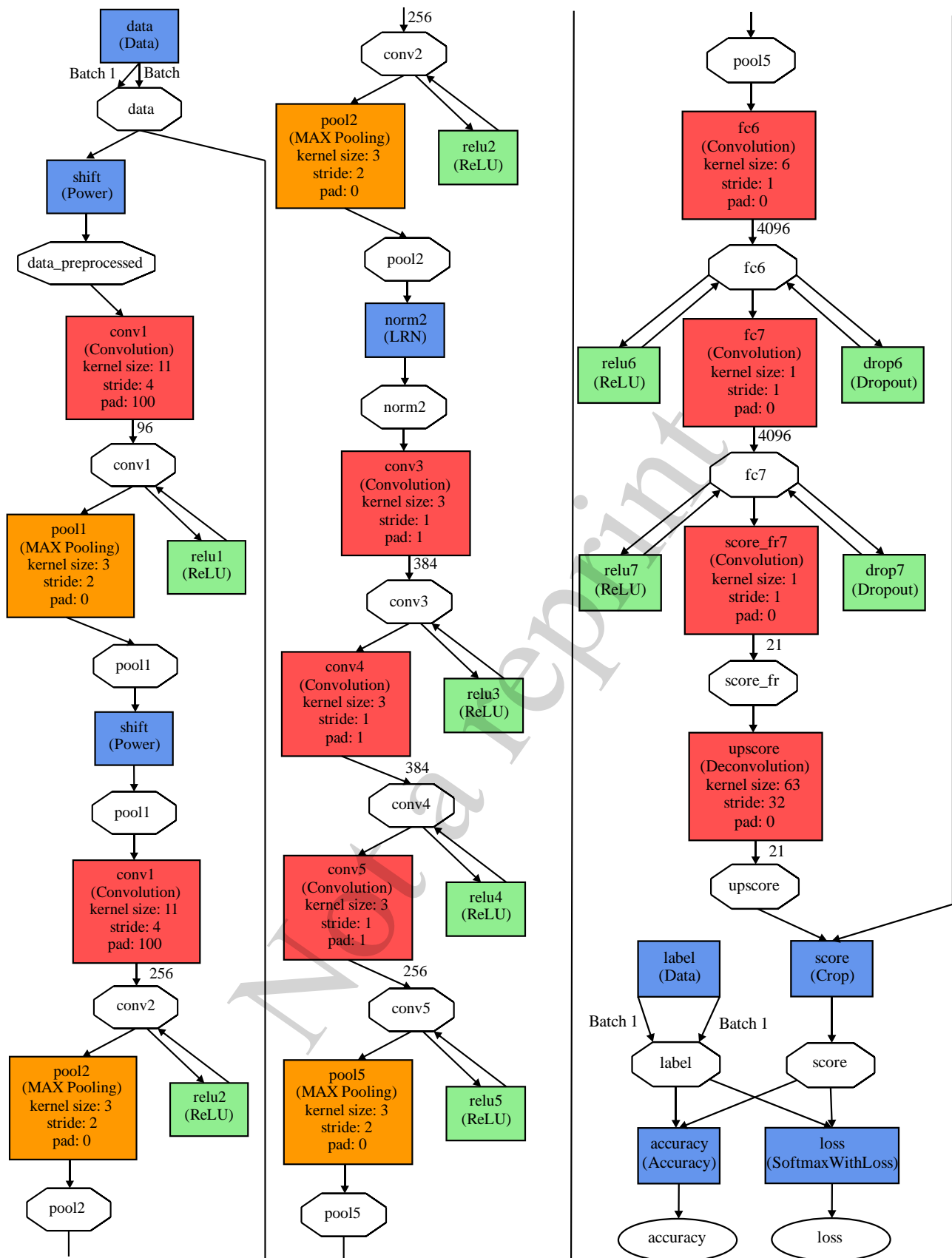


Рис. 1. Архітектура моделі нейронної мережі AlexNet

Таким чином, архітектуру AlexNet було використано як базову для створення спеціалізованої моделі FCN-AlexNet шляхом додавання повнозв'язного згорткового шару через внесення таких змін в DIGITS:

- навчальна та перевірна мережі повинні бути об'єднані в єдину нейромережу;
- спеціалізовані шари для отримання даних із зображення для сегментації SBDDSegDataLayer та VOCSegDataLayer повинні бути замінені простими шарами для отримання даних з LMDB;
- шар power повинен бути доданий до зсувного шару вхідних даних;
- до шару upscore потрібно додати білінійну фільтрацію ваги нейронів;
- потрібно додати шар точності для оцінки ефективності моделі на перевірному наборі зображень;
- процеси навчання та перевірки повинні виконуватися з нормалізацією шляхом введення параметра batch size=1 [15].

Вибір алгоритму оптимізації моделі FCN-AlexNet визначається особливостями сегментації зображення, для якої необхідно мати хорошу збіжність алгоритму, а для практичного використання – високу швидкість. Порівняння алгоритмів [16] показує, що для задачі сегментації зображень найкращі результати швидкодії показують Adam (збільшення на 10–50 %) та стохастичний градієнтний спуск (збільшення на 5–20 %) в залежності від метрики. Також ці алгоритми демонструють хорошу збіжність, особливо стохастичний градієнтний спуск.

Основними показниками ефективності навчання нейромережі, що визначалися в процесі дослідження, було обрано точність і похибки. Точність рахується як відсоток правильно визначених класів (або пікселів, що належать до певного класу) на зображенні відносно всіх класів (або всіх пікселів). Похибка рахується як відсоток неправильно розпізнаних класів (або пікселів, що належать до певного класу) на зображенні відносно всіх класів (або всіх пікселів). Для оцінки ефективності навчання нейромережі визначаються оптимальні параметри нейронної мережі. Цими параметрами є тривалість навчання (число епох), алгоритм оптимізації (адаптивна миттєва оцінка (Adam), стохастичний градієнтний спуск (SGD)), вид зміни швидкості навчання, коефіцієнт gamma, швидкість навчання (величина кроку). Комбінації параметрів в процесі навчання 4-х моделей зведено у табл. 2.

Похибка досліджувалась на навчальній та перевірній вибірках, які являють собою вищенаведені набори розмічених сегментованих зображень (PASCAL VOC та NVIDIA-Aerial Drone). Набір PASCAL VOC містить 20 категорій об'єктів, які включають 1464 зображень для навчання та 1449 – для перевірки. Для набору NVIDIA-Aerial Drone виконано розподіл зображень у співвідношенні 80 % на навчальну вибірку та 20 % на перевірну з використанням кросс-перевірки для оцінки точності моделі.

Значення похибки на навчальній та, особливо, на перевірній вибірці повинно поступово зменшуватись. Це буде свідчити про правильність навчання нейронної мережі та відсутність перенавчання, тобто про адекватність моделі. Валідація результатів навчання може визначатись як поступове зниження похибки на перевірній вибірці. Кількість епох навчання вибирається з умови

отримання найвищої точності на перевіірочній вибірці при відсутності суттєвих коливань числових значень. Критерієм доцільності збільшення епох навчання є поступове підвищення точності на перевіірочній вибірці. Початок падіння точності на перевіірочній вибірці є критерієм перенавчання, відсутність якого є умовою валідації моделі.

Таблиця 2
Комбінації параметрів в процесі навчання

Мо- дель	Тривалість навчання (число епох)	Алгоритм оптиміза- ції	Базуєть- ся на моделі	Вид зміни швидкості навчання, коефіцієнт gamma	Швид- кість на- вчання
Voc-1	25	Adam	FCN- AlexNet	Експоненційний, 0,99	0,0001
Voc-2	30	SGD	FCN- AlexNet	Експоненційний, 0,95	0,0001
Voc-3	50	SGD	FCN- AlexNet	Ступінчастий (розмір кроку=33), 0,1	0,0001
Voc-4	50	SGD	FCN- AlexNet	Експоненційний, 0,97	0,00001

5. Результати дослідження сегментації зображень за допомогою згорткових нейронних мереж

5.1. Дослідження моделей нейромереж для сегментації зображень з набору PASCAL VOC

Дослідження проводилось на попиксельно розмічених зображеннях з набору PASCAL VOC. Оскільки розміри зображень в наборі різні, то у разі потреби перевірки моделі на новому зображенні необхідно виконати попиксельну розмітку сегментованих областей. Крім того, розрахунок точності відбувається в режимі м'якого реального часу, що вимагає від нейромережі високої швидкодії при обмеженому обсягу пам'яті для забезпечення високої точності. Зображення на вхід нейромережі надходить із колірною палітрою RGB (256 кольорів) розміром не більше 10 Мб, а на виході отримується в форматі PNG (lossless). Розміром пакета (batch size) складає 32 з кількістю потоків, рівною 4. Попиксельна розмітка (за наявності) повинна відповідати формату анотації Lightning Memory-Mapped Database. Моделі імпортуються у форматі prototxt. Час виконання сегментації не повинен перевищувати 50 мс для зображення розміру Full HD.

Перевірка ефективності моделей проілюстровано графіками, побудова яких виконувалась автоматично в програмному середовищі DIGITS на основі заданих параметрів, які наведені в табл. 2. Для апаратного прискорення навчання використовувалось середовище Caffe. Кількість значень точності рівний кількості епох навчання. Остаточна точність моделі є точність навчання на останній епосі.

На рис. 2 показано графік перевірки ефективності моделі Voc-1.

З рис. 2 видно, що на навчальній вибірці зі збільшенням кількості етапів навчання (епох) похибка зменшується у абсолютному значенні з 3 до 1,5–2,5, а

після 10-ї епохи стабілізується в діапазоні 1,3–2,7. Точність на перевірочній вибірці досягає 72 % і майже не збільшується після першої епохи.

На рис. 3 показано графік перевірки ефективності моделі Voc-2.

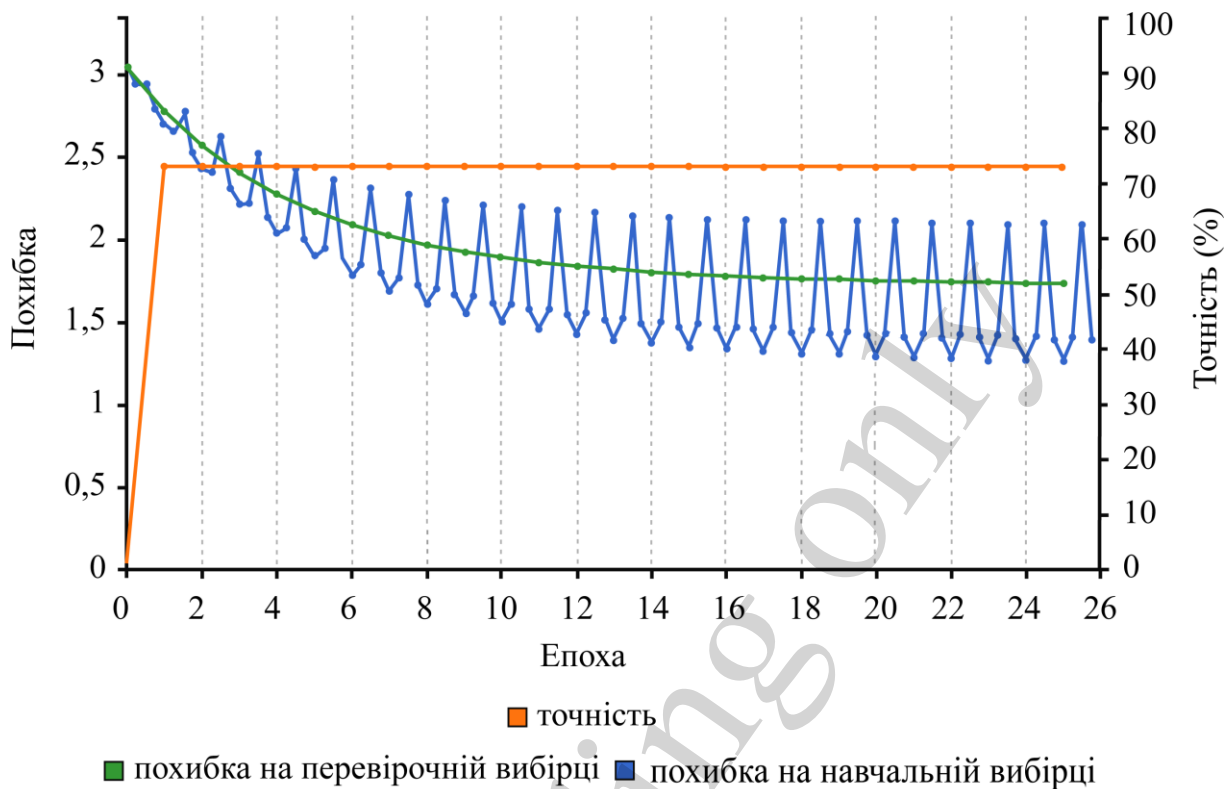


Рис. 2. Графіки зміни точності, похибки на перевірочній вибірці та похибки на навчальній вибірці залежно від епохи для моделі Voc-1

На рис. 3 видно, що на навчальній вибірці зі збільшенням кількості етапів навчання (епох) похибка зменшується у абсолютному значенні з 3 до 0,7–0,2 а після першої епохи і в подальшому практично не змінюється. Точність на перевірочній вибірці досягає 80 % і плавно збільшується після першої епохи до значення 82 %.

На рис. 4 показано графік перевірки ефективності моделі Voc-3.

На рис. 4 видно, що на навчальній вибірці зі збільшенням кількості етапів навчання (епох) похибка зменшується у абсолютному значенні з 3 до 0,6–0,2, а після 10-ї епохи стабілізується в діапазоні 0,4–0,2. Точність на перевірочній вибірці досягає 78 % і плавно збільшується після першої епохи до значення 83 %.

На рис. 5 показано графік перевірки ефективності моделі Voc-4.

На рис. 5 видно, що на навчальній вибірці зі збільшенням кількості етапів навчання (епох) похибка зменшується у абсолютному значенні з 3 до 1,7–0,3, а після 10-ї епохи стабілізується в діапазоні 0,7–0,3. Точність на перевірочній вибірці досягає 75 % і плавно збільшується після першої епохи до значення 81 %.

В результаті досліджень отримані результати перевірки ефективності 4-х моделей занесено в табл. 3.

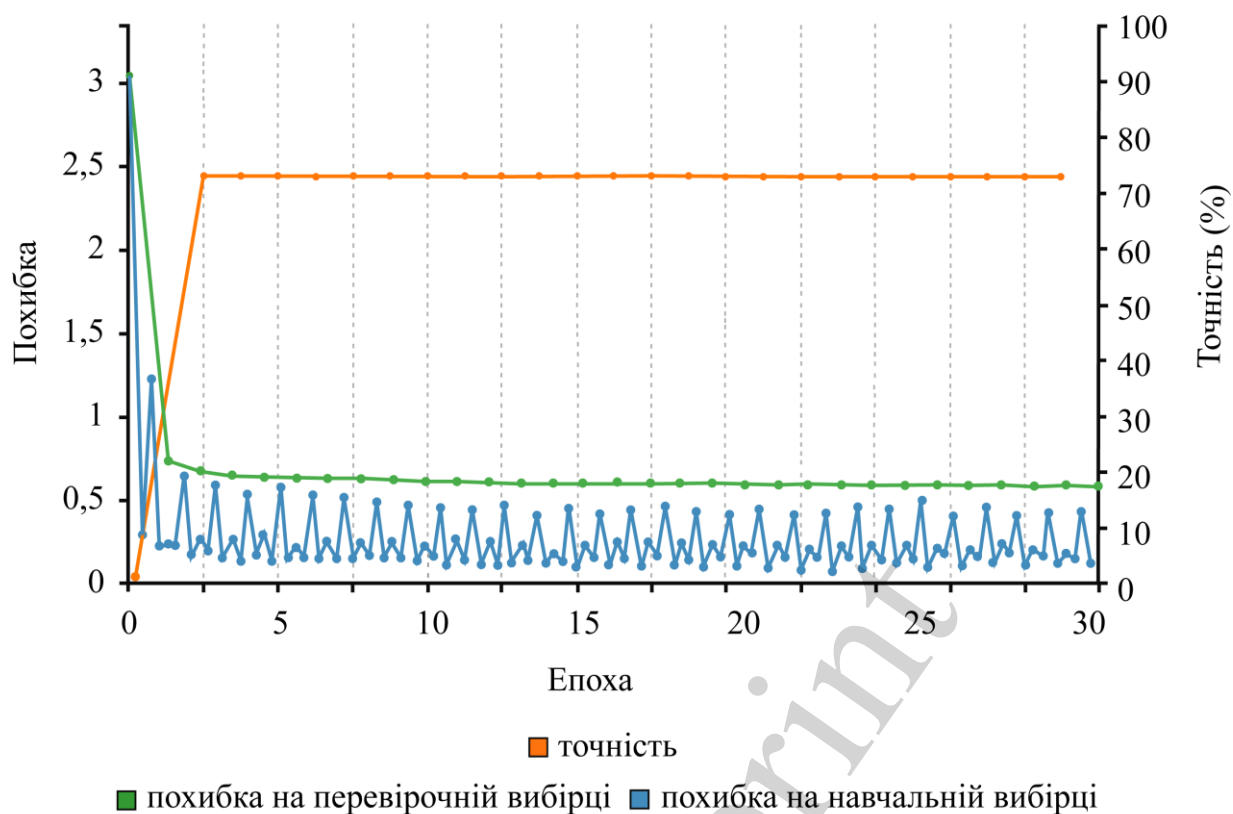


Рис. 3. Графіки зміни точності, похибки на перевірочній вибірці та похибки на навчальній вибірці залежно від епохи для моделі Voc-2

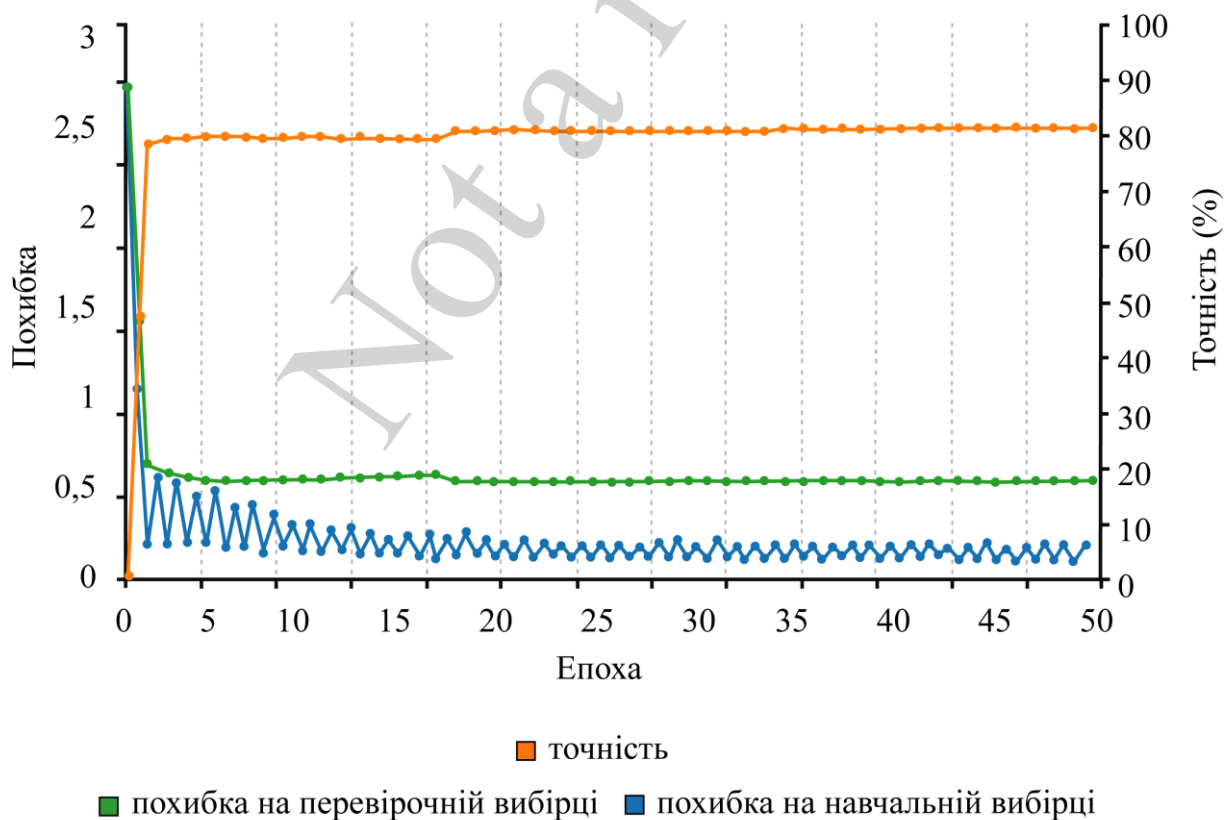


Рис. 4. Графіки зміни точності, похибки на перевірочній вибірці та похибки на навчальній вибірці залежно від епохи для моделі Voc-3

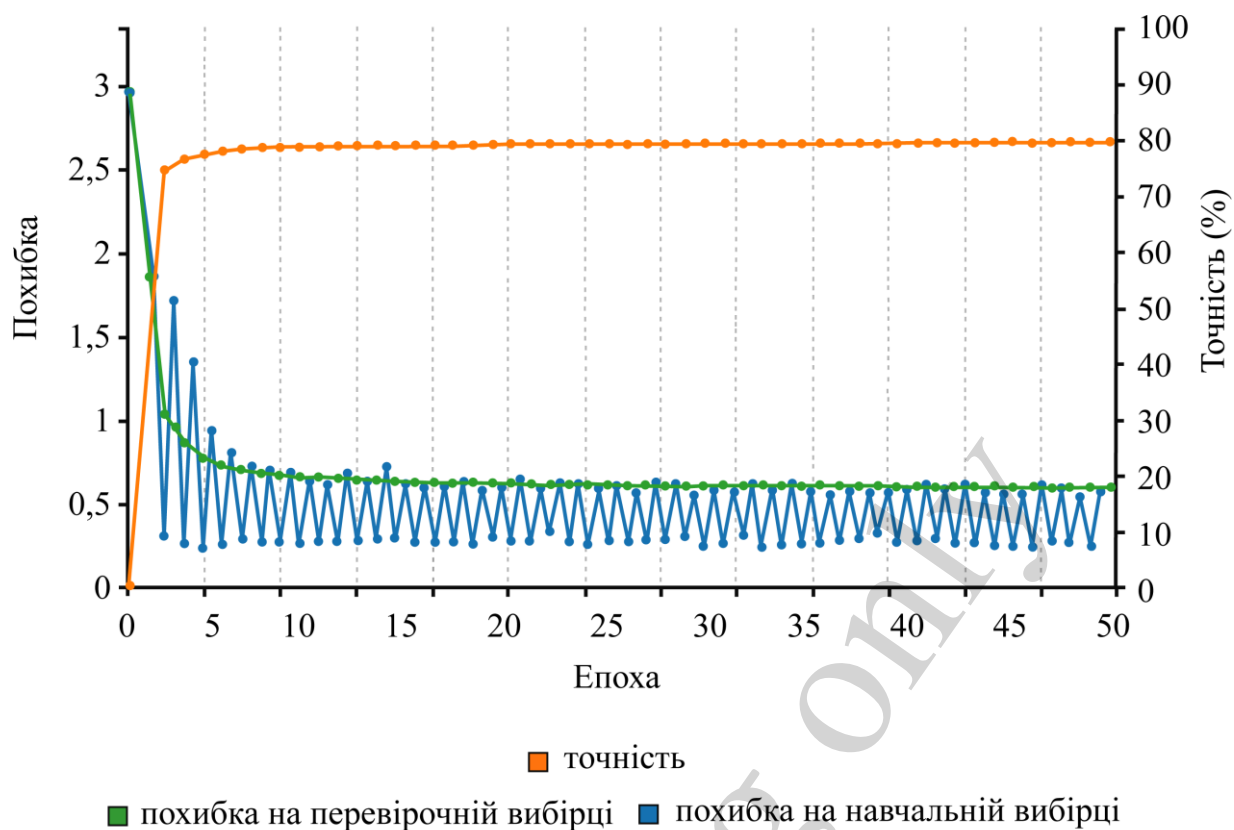


Рис. 5. Графіки зміни точності, похибки на перевіірчній вибірці та похибки на навчальній вибірці залежно від епохи для моделі Voc-4

Таблиця 3

Результати досліджень ефективності моделей із різними параметрами

Назва моделі	Епоха з найкращим результатом/кількість епох	Точність
Voc-1	25/25	72 %
Voc-2	30/30	82 %
Voc-3	50/50	83 %
Voc-4	50/50	81 %

З табл. 3 видно, що найбільшу точність демонструє модель Voc-3 з 83 % при швидкості навчання 0,0001 на основі SGD при ступінчастому способі зміни швидкості навчання. Найменше значення точності складає 72 % для моделі Voc-1, яка використовує алгоритм адаптивної миттєвої оцінки.

На рис. 6 показано сегментацію зображення фотографії тварин в програмному середовищі DIGITS із використанням моделі Voc-3 для навчання при різних епохах. При цьому для апаратного прискорення навчання використовувалось середовище Caffe.

На рис. 6 при виконанні розпізнавання об'єктів виділено 2 області: ■ – «тварина», ■ – «фон».



a



б



в



г

Рис. 6. Сегментації зображення велосипедиста із використанням навченої моделі Voc-3 для різної кількості навчальних епох: *a* – 3 епохи; *б* – 12 епох; *в* – 30 епох; *г* – 50 епох

5. 2. Оцінка навченої моделі Voc-3 для сегментації зображень з набору NVIDIA-Aerial Drone

На практиці сегментація є складовою моніторингу навколишнього середовища за допомогою БПЛА, що вимагає високої точності сегментації для керування та орієнтації в просторі. Відеокамери з БПЛА отримують контрастне зображення, на відміну від зображень в наборі PASCAL VOC. Тому модель з найбільшою точністю Voc-3 потрібно донавчати на іншому наборі зображень, знятих з відеокамери БПЛА. Для вирішення цього завдання використано набір NVIDIA-Aerial Drone. При цьому від моделі вимагається висока швидкодія при обмеженому обсягу пам'яті для забезпечення високої точності в режимі реального часу.

Для перевірки ефективності моделі Voc-3, навченої за допомогою DIGITS, був використаний набір зображень NVIDIA-Aerial Drone. Для апаратного прискорення навчання використовувалось середовище Caffe. В процесі виконання досліджень використовувались значення параметрів, визначені як оптимальні на основі дослідження (табл. 2):

- швидкість навчання – 0,0001;
- тривалість навчання (число епох) – 50;
- алгоритм оптимізації – SGD;
- вид зміни швидкості навчання – ступінчастий;
- коефіцієнт γ – 0,1;
- попередньо навчена модель – повнозв’язна згорткова нейромережа FCN-AlexNet.

Таким чином, отримано модель Drone-1, результати перевірки якої показано на рис. 7.

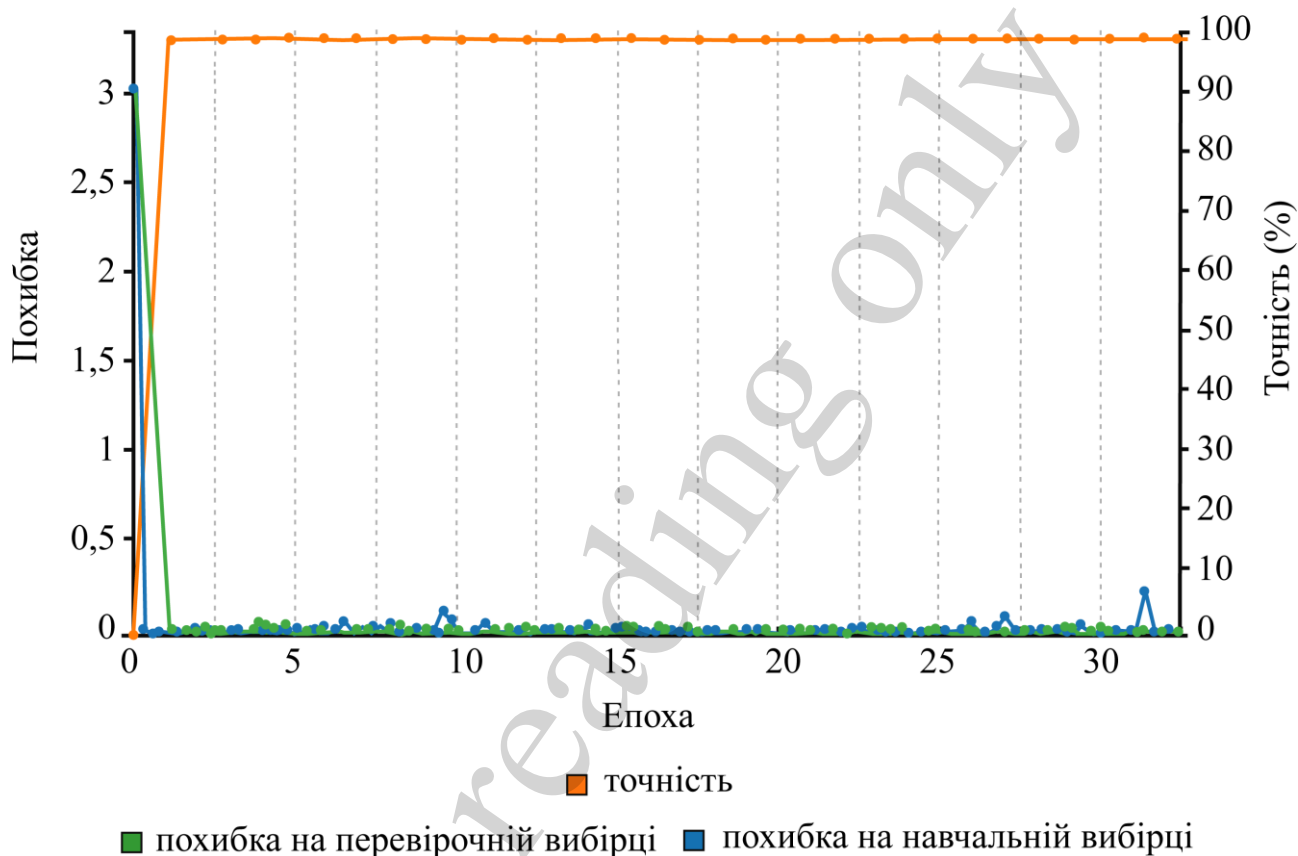


Рис. 7. Графіки зміни точності, похибки на перевірочній вибірці та похибки на навчальній вибірці залежно від епохи для моделі Drone-1

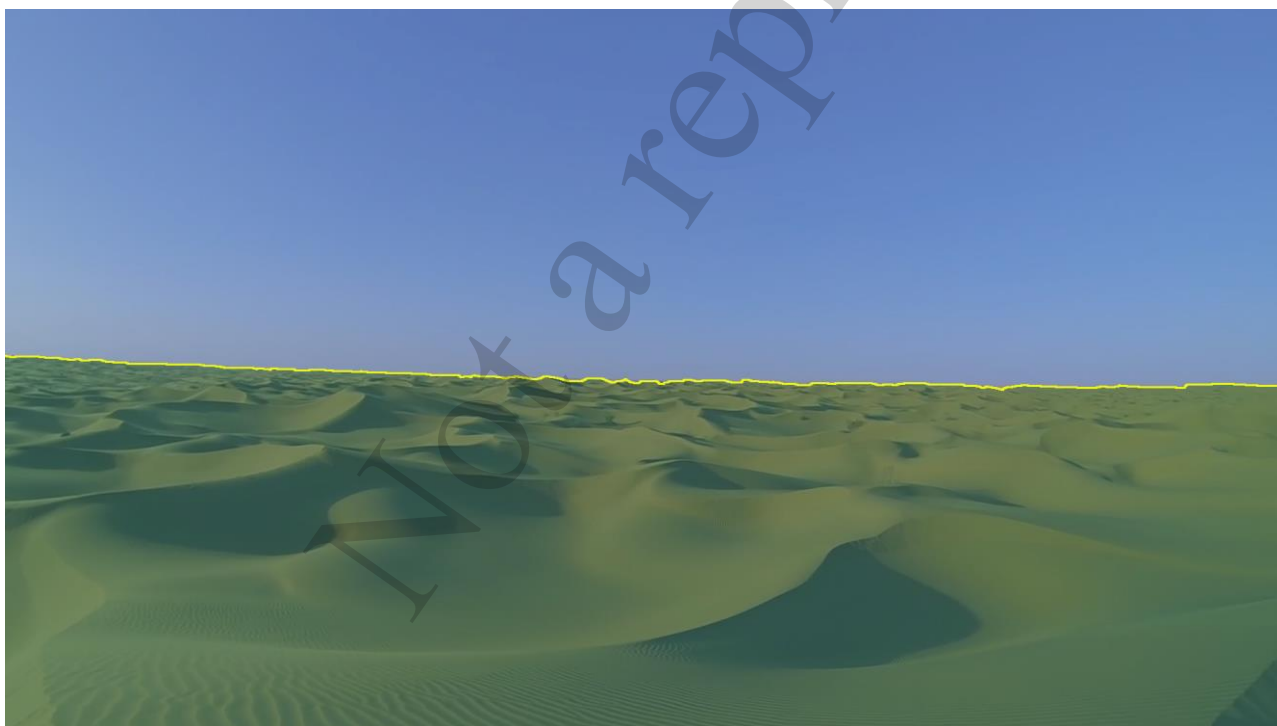
З рис. 7 видно, що похибка в абсолютному значенні для навчальної вибірки зображень стає близькою до 0 після першої епохи навчання, крім епізодичних випадків на певних навчальних епохах. На перевірочній вибірці зображень похибка після першої епохи теж стає близькою до 0, а точність роботи моделі стає близькою до 100 % і практично не змінюється. Після завершення 30-ї епохи значення точності складає 99 %.

На рис. 8 показано сегментація зображення, отриманого з відеокамери БПЛА для моделі Drone-1 в програмному середовищі DIGITS.

На рис. 8 при виконанні розпізнавання об'єктів виділено 2 області: ■ – «земля», ■ – «небо».



a



б

Рис. 8. Сегментація зображення, отриманого з відеокамери БПЛА набору NVIDIA-Aerial Drone для моделі Drone-1: *a* – початкове зображення; *б* – виконана сегментація зображення

Наявні значення швидкодії дозволяють порівняти модель Drone-1 з іншими [2–7], проте є сенс проводити порівняння з близькими по архітектурі моделями, які отримані при навчанні на аналогічній базі. Тому порівняння проведено по

критерію точності з деякими відомими моделями на основі AlexNet, розробленими за схожими параметрами, навченими на основі зображень, отриманих з камер БПЛА. Такими моделями було обрано FIRECAMP2 SLIC, VEDAI, NZAM/ONERA Christchurch, ISPRS Potsdam [17, 18]. Результати оцінки точності моделей занесено в табл. 4.

Таблиця 4
Результати оцінки точності моделей

Назва моделі	Точність, %
FIRECAMP2 SLIC	95
VEDAI	96
NZAM/ONERA Christchurch	97
ISPRS Potsdam	97
Drone-1	98

З табл. 4 видно, що найбільшу точність, а саме 98 %, демонструє модель Drone-1, оскільки інші навчалися на основі зображень, отриманих з камер БПЛА, які не входили в набір NVIDIA-Aerial Drone.

Виконано оцінку адекватності, достовірності та збіжності моделі Drone-1 у порівнянні з іншими [17, 18]. Для цього проведено сегментацію 100 попиксельно розмічених вручну зображень із набору Aerialscapes [19]. Розрахунки виконувались в середовищі Jupiter Notebook на мові Python. Усереднені результати зведені до табл. 5.

Таблиця 5
Результати перевірки для оцінки достовірності, адекватності та збіжності моделі

Назва моделі	Точність, %	Частка зображень із хибно розміченими класами, %	Частка хибно розмічених пікселів, %
FIRECAMP2 SLIC	84	6	12
VEDAI	82	8	17
NZAM/ONERA Christchurch	81	10	20
ISPRS Potsdam	85	4	16
Drone-1	87	5	13

З табл. 5 видно, що найбільшу точність, а саме 87 %, демонструє розроблена модель Drone-1. Це свідчить про високу достовірність моделі Drone-1.

Об'єкти на зображеннях для перевірки розмічені на певні класи, згідно яких відбувається сегментування зображення моделлю, яка виконує позначення класів для виділених областей. При експериментальній перевірці моделей були хибно позитивні (наявність певного класу на зображенні за його відсутності) та хибно негативні (відсутність певного класу на зображенні за його присутності) результати при позначенні класів. З табл. 5 видно, що частка зображень із хибно розміче-

ними класами складає 5 %. Таким чином, частка зображень із правильно розміченими класами складає 95 %, що свідчить про високу адекватність моделі.

Сегментація передбачає врахування частки хибно розмічених пікселів. При експериментальній перевірці моделей були хибно позитивні (позначення пікселя певним класом за відсутності належності до нього) та хибно негативні (непозначення пікселя певним класом при належності до нього) результати розмічення пікселів. З табл. 5 видно, що частка хибно розмічених пікселів складає 13 %. Таким чином, частка правильно розмічених пікселів складає 87 %, що свідчить про високу збіжність моделі.

6. Обговорення результатів дослідження сегментації зображень за допомогою згорткових нейронних мереж

Результати досліджень показують, що навчена модель Drone-1 демонструє високу точність сегментації зображень (рис. 7). Це пояснюється вибором оптимальних параметрів нейронної мережі, а також внесенням згорткового шару в стандартну архітектуру нейромережі. Drone-1 базується на Voc-3. Ця модель продемонструвала найкращі значення точності (табл. 3) в процесі навчання на основі набору PASCAL VOC (рис. 4). Це пояснюється вибором ступінчастої функції для зміни швидкості навчання (табл. 1). Рис. 6 візуалізує отриманий результат, що підтверджує отримані дані. Тому саме Voc-3 була обрана для донавчання на основі набору NVIDIA-Aerial Drone із тими ж параметрами. Створену таким чином модель Drone-1 можна використовувати для сегментації реальних зображень (рис. 8), висока точність якої визначає ефективність системи керування БПЛА. Результати порівняння точності сегментації зображення (рис. 8) для Drone-1 та інших подібних моделей занесено в табл. 4. Drone-1 в порівнянні з іншими демонструє високі значення точності сегментації зображень, які не були в навчальній та перевіірчній вибірках, що вказує на адекватність цієї моделі та відсутність перенавчання.

Точність та швидкодія розробленої моделі нейронної мережі Drone-1 мають більші значення, ніж аналогічні, отримані в [17], на 1–2 % та 20–50 %, відповідно. При цьому процес сегментації не вимагає значних обчислювальних ресурсів на етапі використання моделі. Порівняно з [18] дана модель має вищу точність сегментації зображень на 3 %. Це досягнуто за рахунок об'єднання навчальної та перевіірочної мереж, заміни спеціалізованих шарів простими, додаванням шарів power та точності, білінійної фільтрації ваги нейронів та параметра batch size. Достовірність, адекватність та збіжність розробленої моделі нейронної мережі Drone-1 є порівнювана із іншими моделями [17, 18] та не поступається їм.

Оскільки навчання моделі нейромережі проводилось для зображень з набору NVIDIA-Aerial Drone, то високі значення точності сегментації притаманні зображенням, отриманим із відеокамери дрона, зазвичай через високу контрастність груп пікселів. Для інших видів зображень, імовірно, точність не буде настільки високою. Це потребує додаткових досліджень.

До недоліків варто віднести тривалий процес навчання нейромережі та затратність ресурсів на початковому етапі навчання. Подолання цього недоліку

можливо за рахунок розпаралелювання обчислень на графічних процесора, створення нових компактних архітектур та появи більшої кількості попередньо навчених нейронних моделей.

Розвиток дослідження даної моделі може полягати у подальшому підвищенні точності, швидкодії та зменшенні обсягів обчислювальних ресурсів. Це вимагатиме складного математичного моделювання, врахування предметної області застосування та розробки програмних модулів для конкретної системи.

7. Висновки

1. Досліджено моделі Voc-1, Voc-2, Voc-3, Voc-4 нейромереж на основі набору PASCAL VOC. Встановлено, що найбільшу точність демонструє модель Voc-3 з 83 % при швидкості навчання 0,0001 на основі SGD при ступінчастому способі зміни швидкості навчання. Найменше значення точності складає 72 % для моделі Voc-1, яка використовує алгоритм адаптивної миттєвої оцінки. Це означає, що SGD справляється краще, ніж Adam, оскільки Voc-2 та Voc-4 володіють відчутно більшими показниками точності моделі, а саме 82 % та 81 %, відповідно.

2. Отримано значення точності моделі Drone-1, створеної на основі попередньо навченої моделі Voc-3 для сегментації зображень з набору NVIDIA-Aerial Drone з визначеними на основі дослідження параметрами. Ці значення є досить високими, адже похибка в абсолютному значенні для навчальної вибірки зображень стає близькою до 0 після першої епохи навчання. На перевіірочній вибірці зображень похибка після першої епохи теж стає близькою до 0, а точність роботи моделі стає близькою до 100 % і практично не змінюється. Після завершення 30-ї епохи значення точності складає 99 %. Отримані значення точності дозволяють стверджувати про правильність вибору архітектури мережі та підбору параметрів. Це дає можливість використовувати дану модель для практичних задач сегментації зображень, наприклад, у випадку оцінки величини пожежі, аналізу вегативності поля, класифікації посівів тощо.

Література

1. Білинський, Й. Й., Книш, Б. П., Кулик, Я. А. (2017). Методика оцінювання якості роботи фільтрів приглушення шумів в пакеті MathCAD. Вісник Хмельницького національного університету, 3, 125–130. URL: <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/23238/47857.pdf?sequence=2&isAllowed=y>
2. Kurugollu, F., Sankur, B., Harmanci, A. E. (2001). Color image segmentation using histogram multithresholding and fusion. Image and Vision Computing, 19 (13), 915–928. doi: [https://doi.org/10.1016/s0262-8856\(01\)00052-x](https://doi.org/10.1016/s0262-8856(01)00052-x)
3. Wang, H., Oliensis, J. (2010). Generalizing edge detection to contour detection for image segmentation. Computer Vision and Image Understanding, 114 (7), 731–744. doi: <https://doi.org/10.1016/j.cviu.2010.02.001>
4. Felzenszwalb, P. F., Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 59 (2), 167–181. doi: <https://doi.org/10.1023/b:visi.0000022288.19776.77>

5. Chitade, A. Z., Katiyar, S. K. (2010). Colour based image segmentation using k-means clustering. *International Journal of Engineering Science and Technology*, 2 (10), 5319–5325. URL: https://www.researchgate.net/publication/50361273_Color_based_image_segmentation_using_K-means_clustering
6. Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, 482. URL: <http://people.sabanciuniv.edu/berrin/cs512/lectures/Book-Bishop-Neural%20Networks%20for%20Pattern%20Recognition.pdf>
7. Chaudhuri, D., Agrawal, A. (2010). Split-and-merge Procedure for Image Segmentation using Bimodality Detection Approach. *Defence Science Journal*, 60 (3), 290–301. doi: <https://doi.org/10.14429/dsj.60.356>
8. Keuchel, J., Schnorr, C. (2003). Efficient Graph Cuts for Unsupervised Image Segmentation using Probabilistic Sampling and SVD-based Approximation. University of Mannheim. URL: https://madoc.bib.uni-mannheim.de/1805/1/2003_9.pdf
9. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2015). Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv.org*. URL: <https://arxiv.org/pdf/1412.7062.pdf>
10. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, 1097–1105. URL: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
11. Ситник, В. Ф. (2004). Системи підтримки прийняття рішень. К.: КНЕУ, 614. URL: <http://kist.ntu.edu.ua/textPhD/sppr1.pdf>
12. Shelhamer, E., Long, J., Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (4), 640–651. doi: <https://doi.org/10.1109/tpami.2016.2572683>
13. Semantic Segmentation with SegNet. URL: <https://github.com/dusty-nv/jetson-inference/blob/master/docs/segnet-dataset.md>
14. Saxena, S. (2021). Introduction to The Architecture of Alexnet. *Analytics Vidhya*. URL: <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-the-architecture-of-alexnet/>
15. How to do Semantic Segmentation using Deep learning. URL: <https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning>
16. Kingma, D. P., Ba, J. (2015). Adam: a method for stochastic optimization. *arXiv.org*. URL: <https://arxiv.org/pdf/1412.6980.pdf>
17. Solovey, D. (2017). Terrain Classification from High Resolution Aerial. Images Using Deep Learning. URL: https://fenix.tecnico.ulisboa.pt/downloadFile/1689244997257919/Resumo_174132.pdf
18. Audebert, N., Le Saux, B., Lefèvre, S. (2017). Segment-before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images. *Remote Sensing*, 9 (4), 368. doi: <https://doi.org/10.3390/rs9040368>
19. AEROSCAPES. Aerial Semantic Segmentation Benchmark. URL: <https://github.com/ishann/aeroscapes>